Simultaneous unbiased estimates of multiple downed wood attributes in perpendicular distance sampling

Mark J. Ducey, Michael S. Williams, Jeffrey H. Gove, and Harry T. Valentine

Abstract: Perpendicular distance sampling (PDS) is a fast probability-proportional-to-size method for inventory of downed wood. However, previous development of PDS had limited the method to estimating only one variable (such as volume per hectare, or surface area per hectare) at a time. Here, we develop a general design-unbiased estimator for PDS. We then show how that estimator can be used to develop simple measurement protocols that allow simultaneous, unbiased estimation of multiple downed wood variables, including logs per hectare, length of logs per hectare, surface area or area coverage per hectare, and volume per hectare.

Résumé : L'échantillonnage à distance perpendiculaire (EDP) est une méthode rapide pour faire l'inventaire du bois au sol dont la probabilité est proportionnelle à la dimension. Cependant, cette méthode d'EDP a jusqu'à maintenant été développée pour estimer une seule variable (telle que le volume à l'hectare ou la superficie à l'hectare) à la fois. Ici, nous développons un estimateur non biaisé et général pour l'EDP. Nous montrons ensuite comment cet estimateur peut être utilisé pour développer des protocoles simples de mesure qui permettent d'obtenir simultanément une estimation non biaisée de plusieurs variables du bois au sol, incluant le nombre de billes à l'hectare, la longueur des billes à l'hectare, la superficie ou la surface couverte à l'hectare et le volume à l'hectare.

[Traduit par la Rédaction]

Introduction

The past two decades have seen dramatic growth in the awareness and understanding of the role of dead downed wood (also known as coarse woody material (CWM) and coarse woody debris (CWD); hereinafter refer to as CWM) in a variety of ecosystem states and processes (Harmon et al. 1986; Hagan and Grove 1999; Ståhl et al. 2001) including wildlife habitat, carbon and nutrient sinks and cycles, and wildland fire behavior. This growing interest has been coupled with increased attention to efficient methods for sampling CWD.

One promising set of techniques is perpendicular distance sampling (PDS; Williams and Gove 2003). PDS is a probability proportional to size sampling (PPS) technique; in its original development, sampling was done with probability proportional to volume. As a result few, if any, actual measurements of log size are needed to obtain a design-unbiased estimate of the volume of CWM (whether per unit area or total for a tract). The result is a sampling method that is nearly optimal in terms of its theoretical variance (Williams and Gove 2003) and very fast. Williams et al. (2005a) generalized the method for sampling with probability proportional to

CWM surface area; Williams et al. (2005b) present solutions to practical problems including correction for slope and for curved, forked, or leaning CWM. Simulation studies (Williams and Gove 2003; Williams et al. 2005a) and field trials[2] show that PDS can provide substantial improvements in efficiency over traditional methods such as line intersect sampling (e.g., Brown 1974).

An unfortunate limitation of the previously developed variants of PDS is that only one CWM variable can be estimated at a time. For example, in PDS with probability proportional to volume a simple count of tallied logs multiplied by a volume factor gives an estimate of volume per hectare. However, without additional and seemingly difficult measurements no estimate can be provided for number of logs per hectare or for other CWM variables (Williams et al. 2005a). The situation is similar to that in the most familiar PPS technique in forestry, horizontal point sampling. In horizontal point sampling, a count of tallied trees at a point multiplied by the basal area factor gives an estimate of basal area per hectare. To expand other tree attributes to a per hectare basis, one must also measure the basal area of the trees (or diameter at breast height, which is converted to basal area under an assumption of circularity, giving a

M.J. Ducey.[1] Department of Natural Resources, University of New Hampshire, Durham, NH 03824, USA.
M.S. Williams. Risk Assessment and Residue Division, Food Safety Inspection Service, USDA, 2150 Centre Avenue, Building B, Mail Stop 2E7, Fort Collins, CO 80526, USA.
J.H. Gove and H.T. Valentine. USDA Forest Service, Northeastern Research Station, P.O. Box 640, Durham, NH 03824, USA.

[1]Corresponding author (e-mail: mjducey@cisunix.unh.edu).
[2]Ducey, M.J., Williams, M.S., Roberge, S., Kenning, R., and Gove, J.H. Distance limited perpendicular distance sampling for coarse woody material: theory and field results. Environ. Ecol. Stat. Submitted.

nearly exact approximation for almost all trees; Grosenbaugh 1958). A similarly direct attack on the problem in PDS would require, for example, accurate measurement of the volume of downed logs that are tallied. This task is a forbidding one considering the variety of irregular shapes that downed, decaying logs may take. Exacting measurement of the volume of irregular logs would eliminate any practical advantages, especially in terms of speed, that PDS might have.

Here, we solve the problem of obtaining design-unbiased estimates of multiple CWM variables in PDS using simple, familiar, and quick measurements. First, we develop a general design-unbiased estimator that takes advantage of the deep connection between PDS and importance sampling (Gregoire et al. 1987). We then use the estimator to develop specific measurement techniques and estimators for CWM variables including logs per hectare, length per hectare, surface area coverage, and log volume and biomass per hectare when PDS is used with probability proportional to volume or probability proportional to surface area.

Overview of PDS

We begin with a very brief review of PDS for the convenience of readers who may not be familiar with this new method and to establish notation for later sections. This review is not exhaustive; for greater detail, readers may wish to consult Williams and Gove (2003) and Williams et al. (2005*a*, 2005*b*).

Suppose we are interested in the total of some attribute of pieces of CWM that lie within a defined tract, $\mathcal{A}$, that has area $|\mathcal{A}|$ (ha). For example, we might be interested in V, the total volume of CWM (m^3), or equivalently the volume per unit area $V/|\mathcal{A}|$ (m^3/ha). There are N logs lying in $\mathcal{A}$, but we do not know N. We will sample the population by establishing one or more sample points uniformly at random within $\mathcal{A}$. Without loss of generality, we initially consider a single sample point.

From the sample point, which plays the same "plot center" role as a sample point in horizontal point sampling, we scan for all logs that may qualify to be tallied. First, we determine whether a log has a "perpendicular point," that is, a point on the log (or the axis connecting the ends of the log; Williams et al. 2005*b*) where the line of sight from the sample point is perpendicular to the log axis. If not, the log is disregarded. If so, then the distance from the sample point to the perpendicular point is compared with a critical or limiting distance that depends on some attribute of the log taken at the perpendicular point. Specifically, let D_i be the distance to the perpendicular point on the ith log. The log will be tallied if $D_i \leq D_L$ where

[1] $\qquad D_L = kx_i(h)$

Here, $x_i(h)$ is the value of some attribute x on the ith log evaluated at the perpendicular point, which is located at a distance h (m) from the basal end of the log. Let the total length of the ith log be H_i. The inclusion zone for the ith log is the zone in which sample points can fall and the log will be tallied. Let the inclusion zone be a_i and its area be denoted as $|a_i|$ (m^2). Then

Fig. 1. Overview of perpendicular distance sampling (PDS). (*a*) A sample point (*) has fallen in the vicinity of four logs. The observer sights on the perpendicular point of each log; the perpendicular line of sight is shown as a broken line. Note that there is no perpendicular point for log 4 because of its position and orientation; log 4 is ignored. (*b*) "Inclusion zones" for the four logs. In this case, because the sample point has fallen within the inclusion zones of logs 1 and 2, they are tallied; logs 3 and 4 are not.
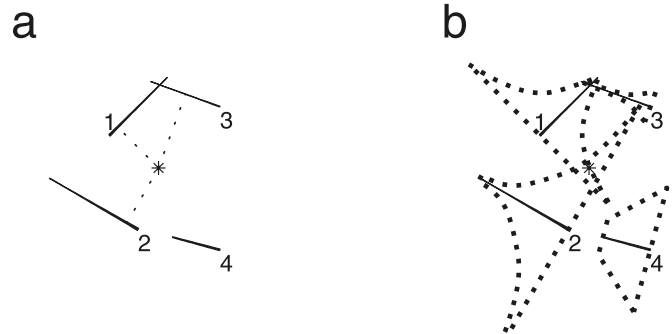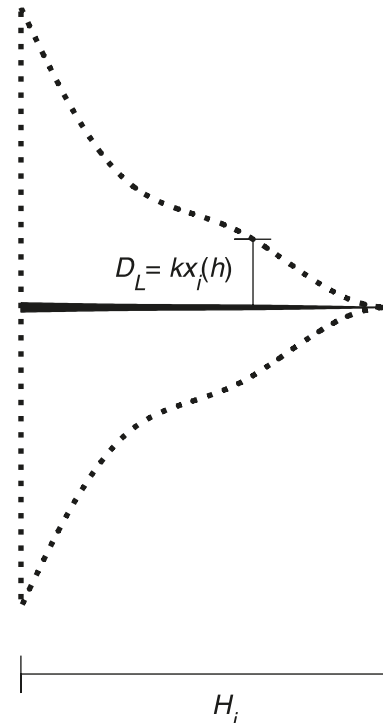

Fig. 2. Expanded view of the inclusion zone for an individual log in perpendicular distance sampling (PDS). H_i is the length of the ith log axis in the horizontal plane and $x_i(h)$ is an attribute of the ith log taken perpendicular to the log axis at point h. The choice of the attribute x determines the "size" proportional to which PDS samples. The choice of the constant k determines the effective "factor" (e.g., volume factor, area factor) in PDS.

[2] $\qquad |a_i| = 2 \int_0^{H_i} kx_i(h)\mathrm{d}h$

The field procedure and corresponding inclusion zones are illustrated in Fig. 1. An expanded view of the inclusion zone of a single log is shown in Fig. 2.

The design of a PDS method involves two choices. The

first is the selection of an attribute to use as $x_i(h)$. PDS is a PPS method; the choice of $x_i(h)$ determines the meaning of "size." For example, if $x_i(h)$ is the cross-sectional area of the log $x_{a,i}(h)$, then PDS samples are taken with probability proportional to the volume v_i of logs since

$$v_i = \int_0^{H_i} x_{a,i}(h)\mathrm{d}h$$

so

$$
\begin{aligned}
|a_i| &= 2\int_0^{H_i} k x_{a,i}(h)\mathrm{d}h \\
&= 2k v_i
\end{aligned}
$$

The second choice, the value of k, determines the effective "volume factor" of the sampling approach. The probability that the ith log will be tallied, or π_i in a standard Horvitz–Thompson estimator (Horvitz and Thompson 1952), is $\pi_i = |a_i|/[(10\,000\text{ m}^2/\text{ha})|\mathcal{A}|]$. Let δ_i be an indicator variable; δ_i is equal to 1 if the ith log is tallied and δ_i is equal to 0 otherwise. The corresponding estimator for volume per hectare is

$$
\begin{aligned}
\frac{\widehat{V}}{|\mathcal{A}|} &= \frac{1}{|\mathcal{A}|}\sum_{i=1}^{N}\delta_i\frac{v_i}{\pi_i} \\
&= \frac{1}{|\mathcal{A}|}\sum_{i=1}^{N}\delta_i\frac{v_i}{\dfrac{2k v_i}{10\,000|\mathcal{A}|}} \\
&= \sum_{i=1}^{N}\delta_i\frac{10\,000}{2k} \\
&= \frac{10\,000n}{2k}
\end{aligned}
$$

where n is the number of logs tallied and $10\,000/2k$ is the volume factor (Williams and Gove 2003).

Development of PDS for sampling surface area or ground area coverage is similarly straightforward. When $x_i(h)$ is log circumference, PDS samples with probability proportional to surface area (or almost exactly so; Williams et al. 2005a). When $x_i(h)$ is the width of the log perpendicular to the log axis and parallel to the horizontal plane, PDS samples with probability exactly proportional to horizontal ground surface coverage.

We emphasize that in the original development of PDS, few actual log measurements were needed because it was only necessary to determine whether a log was close enough to the sample point to be tallied. In practice, the ability of experienced foresters to estimate diameters and distances coupled with a simple chart allows rapid and accurate determination of the inclusion or exclusion of most logs without direct measurement. Measurement is only needed when a log is "borderline." The situation is directly analogous to horizontal point sampling when basal area is the variable of interest (although a chart is used in PDS rather than a physical gauge of some sort). The result is an extremely rapid field technique.[2]

PDS (and PPS sampling in general) provides a "size factor" because the inclusion zone area is proportional to size, so that size cancels out in the standard Horvitz–Thompson estimator (Horvitz and Thompson 1952) when size is also in the numerator. However, when other variables are inserted in the numerator, size does not necessarily cancel out. To use the Horvitz–Thompson estimator for other variables, we would have to measure the size of downed logs accurately, where size might be volume, surface area, or ground surface coverage. This task can be difficult, if not impossible, for real logs in the field. We need another approach.

General design-unbiased estimator

As an alternative to the Horvitz–Thompson estimator (Horvitz and Thompson 1952) we consider an estimator motivated by the connection between PDS and importance sampling (Gregoire et al. 1987). PDS is a direct physical implementation of importance sampling via von Neumann's (1951) acceptance–rejection method, in which $\mathcal{A}$ is the region within which an initial uniform random deviate is generated. For the purpose of considering the ith log in the population, let us assign this uniform random deviate Cartesian coordinates that depend on the position and orientation of the log. The first Cartesian coordinate is h, and the second is D_i. Let the origin $[0,0]$ be at the basal end of the log axis, and let the distal end fall at $[H_i, 0]$. The deviate is rejected if it falls outside a_i, the shape of which is governed by the value along the h-axis of $x_i(h)$. If $h < 0$ or $h > H_i$ then the deviate will certainly be rejected because the log does not extend into this region (i.e., $x_i(h) = 0$). But, if $0 \leq h \leq H_i$ then the coordinate D_i comes into play. The deviate will be accepted if and only if $-kx_i(h) \leq D_i \leq kx_i(h)$ and rejected otherwise. In other words, $2kx_i(h)$ is an auxiliary function for an importance sample of h from the ith log.

This connection between PDS and importance sampling motivates an alternative estimator, which we develop here as a conjecture and then prove below. Let $Y = \sum_{i=1}^{N}y_i$ where $y_i = \int_0^{H_i} x_i(h)\mathrm{d}h$ be an attribute of a population of N logs in $\mathcal{A}$. We sample using PDS with $D_L = 2kx_i(h)$, which provides a straightforward estimate of Y or of $Y/|\mathcal{A}|$, as shown above. But suppose we are interested in some other quantity $Z = \sum_{i=1}^{N}z_i$ where $z_i = \int_0^{H_i} f_i(h)\mathrm{d}h$. We will show that a general design-unbiased estimator of $Z/|\mathcal{A}|$ is

$$[3] \qquad \frac{\widehat{Z}}{|\mathcal{A}|} = \frac{10\,000}{2k}\sum_{i=1}^{N}\delta_i\frac{f_i(h)}{x_i(h)}$$

where $\delta_i = 1$ if log i is sampled and $\delta_i = 0$ otherwise, as before.

Two equally valid proofs of the design-unbiasedness of the estimator in eq. 3 are available. Here, we take up the proof from a conventional probability sampling perspective. A proof based on a Monte Carlo approach, which is less familiar in forestry literature, is presented in Appendix A.

To prove that eq. 3 is design-unbiased, we first consider the distribution of h given that a log has been sampled. Consider inclusion zones such as those in Fig. 2. If the sample point is located uniformly at random within the inclusion zone then h is not selected uniformly at random on the interval $0 \leq h \leq H_i$, but has an unequal probability distribution. The cumulative distribution function is

$$[4] \quad \begin{aligned} P(h < h') &= \frac{2k \int_0^{h'} x_i(h)\mathrm{d}h}{2k \int_0^{H_i} x_i(h)\mathrm{d}h} \\ &= \frac{2k \int_0^{h'} x_i(h)\mathrm{d}h}{y_i} \end{aligned}$$

Differentiating with respect to h', we obtain the probability density function

$$[5] \quad \begin{aligned} p(h) &= \frac{\mathrm{d}}{\mathrm{d}h'} P(h < h') \\ &= \frac{1}{y_i} \frac{\mathrm{d}}{\mathrm{d}h'} \int_0^{h'} x_i(h)\mathrm{d}h \\ &= \frac{x_i(h)}{y_i} \end{aligned}$$

The next step in the proof is to calculate the expected value of the joint random variable $r_i = \delta_i f_i(h) / x_i(h)$. Now, with probability $(1 - \pi_i)$, $r_i = 0$. So we may write

$$[6] \quad \begin{aligned} \mathrm{E}[r_i] &= \mathrm{E}[r_i | \delta_i = 1] \pi_i \\ &= \pi_i \mathrm{E}_h \left[\frac{f_i(h)}{x_i(h)} \right] \end{aligned}$$

Now, conditional on the event that the ith log is sampled, h is sampled with probability proportional to $x_i(h)$ (eq. 5). So we may rewrite eq. 6 as

$$[7] \quad \begin{aligned} \mathrm{E}[r_i] &= \pi_i \int_0^{H_i} r_i \frac{x_i(h)}{y_i} \mathrm{d}h \\ &= \frac{\pi_i}{y_i} \int_0^{H_i} r_i x_i(h)\mathrm{d}h \end{aligned}$$

Note that if the ith log is sampled then $r_i = f_i(h) / x_i(h)$. So we may further simplify

$$\begin{aligned} \mathrm{E}[r_i] &= \frac{\pi_i}{y_i} \int_0^{H_i} f_i(h)\mathrm{d}h \\ &= \frac{\pi_i z_i}{y_i} \end{aligned}$$

But, accounting for units,

$$\begin{aligned} \pi_i &= \frac{|a_i|}{10\,000|\mathcal{A}|} \\ &= \frac{2k y_i}{10\,000|\mathcal{A}|} \end{aligned}$$

so that

$$[8] \quad \mathrm{E}[r_i] = \frac{2k}{10\,000|\mathcal{A}|} z_i$$

Substituting eq. 8 into eq. 3, we obtain

$$[9] \quad \begin{aligned} \frac{\mathrm{E}[\widehat{Z}]}{|\mathcal{A}|} &= \frac{10\,000}{2k} \sum_{i=1}^{N} \frac{2k}{10\,000|\mathcal{A}|} z_i \\ &= \frac{1}{|\mathcal{A}|} \sum_{i=1}^{N} z_i \\ &= \frac{Z}{|\mathcal{A}|} \end{aligned}$$

which proves the design-unbiasedness of the estimator in eq. 3.

Sampling with probability proportional to volume

It may not be intuitively obvious that the general estimator in eq. 3 leads to field procedures and specific estimators that are simple and practical. To arrive at such procedures and estimators, let us first consider the case in which PDS is used with probability proportional to volume, so that $x_i(h) = x_{a,i}(h)$ is the cross-sectional area of the log taken perpendicular to the horizontal projection of the log axis (Williams and Gove 2003; Williams et al. 2005b). We will consider several candidate variables (Z or z_i) in turn. For simplicity in presenting the estimators, we will use the "volume factor" of the sample, $F_v = 10\,000 / 2k$, which has units m^3/ha.

Logs per hectare

$Z = N$ when estimating logs per hectare, so $z_i = 1$. A simple approach uses $f_i(h) = 1/H_i$, since $\int_0^{H_i} 1/H_i \, \mathrm{d}h = 1$. The summand in the estimator, $f_i(h)/x_i(h)$, is simply $1/[x_{a,i}(h)H_i]$. So the estimator for logs per hectare is

$$\frac{\widehat{N}}{|\mathcal{A}|} = F_v \sum_i \frac{1}{x_{a,i}(h)H_i}$$

Note that for convenience, we have dropped the δ_i term, so the summation is only over the logs that are actually tallied at the sample point. The log length (m) and the log cross-sectional area at the perpendicular point (m^2) are the required measurements for design-unbiased estimation of logs per hectare.

Length (m) per hectare

$L = \sum_{i=1}^{N} H_i$ is naturally estimated by line intersect sampling (cf. Brown 1974; de Vries 1986) as a constant times the number of log intersections. In our context, $z_i = H_i$ and $f_i(h) = 1$, since $\int_0^{H_i} \mathrm{d}h = H_i$. The summand in the estimator $f_i(h)/x_i(h)$ is $1/x_{a,i}(h)$. So the estimator for length of logs per hectare is

$$\frac{\widehat{L}}{|\mathcal{A}|} = F_v \sum_i \frac{1}{x_{a,i}(h)}$$

The log cross-sectional area at the perpendicular point (m^2) is the only required measurement for design-unbiased estimation of log length per hectare.

Surface area (m²) per hectare

Up to a correction factor for taper that is negligible in practice and ignoring the log ends, the surface area, $s_{a,i}$, of a log is $s_{a,i} = \int_0^{H_i} c_i(h)dh$, where $c_i(h)$ is the circumference of the log or the relevant portion of the circumference (Williams et al. 2005a). For example, if the surface area not in contact with the ground is of interest then the relevant portion of the circumference is that part not in contact with the ground. Here, $f_i(h) = c_i(h)$, with $c_i(h)$ measured in metres, and the summand in the estimator $f_i(h)/x_i(h)$ is $c_i(h)/x_{a,i}(h)$. So the estimator for surface area of logs per hectare is

$$\frac{\widehat{S}}{|\mathcal{A}|} = F_v \sum_i \frac{c_i(h)}{x_{a,i}(h)}$$

The best method of measuring the ratio $c_i(h)/x_{a,i}(h)$ will depend on the application. However, one special case deserves further mention: if the log is approximated by a circular cross-section

$$\frac{c_i(h)}{x_{a,i}(h)} = \frac{\pi d_i(h)/100}{\pi/40\,000 d_i(h)^2}$$
$$= \frac{400}{d_i(h)}$$

where $d_i(h)$ is the log diameter at the perpendicular point (cm).

Ground coverage (m²) per hectare

The ground coverage, g_i, of a log (projected onto the surface) is $g_i = \int_0^{H_i} w_i(h)/100 dh$, where $w_i(h)$ is the ground-width of the log (cm) measured perpendicular to the log axis and parallel to the horizontal plane. Here, $f_i(h) = w_i(h)/100$ and the summand in the estimator $f_i(h)/x_i(h)$ is $w_i(h)/100 x_{a,i}(h)$. So the estimator for ground coverage of logs per hectare is

$$\frac{\widehat{G}}{|\mathcal{A}|} = F_v \sum_i \frac{w_i(h)/100}{x_{a,i}(h)}$$

Again, the best method of measuring the ratio $[w_i(h)/100]/[x_{a,i}(h)]$ will depend on the application and may even depend on the characteristics of the log. If the log is well-approximated by an elliptical cross-section

$$\frac{w_i(h)/100}{x_{a,i}(h)} = \frac{w_i(h)/100}{\pi/40\,000 w_i(h) d_{v,i}(h)}$$
$$= \frac{400}{\pi d_{v,i}(h)}$$

where $d_{v,i}(h)$ is the log "diameter" measured in the "vertical direction" (cm). If the log is approximately circular, then $w_i(h) \approx d_{v,i}(h) \approx d_i(h)$, so

$$\frac{w_i(h)}{x_{a,i}(h)} = \frac{400}{\pi d_i(h)}$$

Finally, for highly irregular logs, note that the fraction $[w_i(h)/100]/[x_{a,i}(h)]$ is just the reciprocal of the mean depth of the log (m), measured along $w_i(h)$. For heavily decayed logs, this measurement can be obtained by systematic or random sampling across the log, using a sharpened metal rod as a probe.

Log biomass (kg) per hectare

The biomass, b_i, of a log is $b_i = \int_0^{H_i} \rho_i(h) x_{a,i}(h)dh$, where $\rho_i(h)$ is the mean dry density (kg/m³) of an infinitesimal "cookie" perpendicular to the log axis at h. For practical purposes, the fraction $[f_i(h)/x_i(h)] = \rho_i(h)$ will be well-approximated by the dry mass to volume ratio (kg/m³) of a reasonably finitely thin sample cookie cut at h. The corresponding estimator is

$$\frac{\widehat{B}}{|\mathcal{A}|} = F_v \sum_i \rho_i(h)$$

If such a cookie is too large for transportation or drying, estimation of the dry mass to volume ratio can be easily turned into a straightforward subsampling problem. For example, plugs might be centered on points on the cookie surface using simple random, systematic, or stratified sampling and removed with a tenon cutter for later analysis (including volume and dry mass determination) in the laboratory. The appropriate estimator for obtaining $\rho_i(h)$ from the plug samples would depend on the specific sampling approach used.

Sampling with probability proportional to ground coverage

Sampling with probability proportional to surface area (Williams et al. 2005a) or ground coverage changes the variable in the denominator of the estimator. In this section, we will present the estimators associated with a range of variables when ground coverage is the "size" of the log. Comparison of the estimators in this section with those in the previous section should amply illustrate the procedure for constructing estimators when other variables (e.g., surface area) are used as log size.

Sampling with probability proportional to ground coverage is of special interest in practical sampling. The only variable needed to check whether "borderline" logs are included in the sample is the horizontal "diameter," $w_i(h)$, which can almost always be measured simply and exactly using calipers. Furthermore, sampling with probability proportional to surface area ($D_L \propto w_i(h)$) avoids the problem of "runaway" limiting distances that can occur in sampling with probability proportional to volume ($D_L \propto x_{a,i}(h) \propto d_i(h)^2$).

As in the previous section, we simplify the presentation of the estimators by using the "ground coverage factor," $F_G = 10\,000/2k$, which has units m²/ha.

Logs per hectare

As before, $Z = N$ when estimating logs per hectare, so $z_i = 1$ and $f_i(h) = 1/H_i$. The summand in the estimator, $f_i(h)/x_i(h)$, is now $100/[w_i(h)H_i]$ where the 100 in the numerator converts $w_i(h)$ from centimetres to metres. So the estimator for logs per hectare is

$$\frac{\widehat{N}}{|\mathcal{A}|} = F_G \sum_i \frac{100}{w_i(h) H_i}$$

The log length (m) and the log ground-width at the perpendicular point (cm), which are both reasonably easy to measure accurately, are the required measurements for design-unbiased estimation of logs per hectare.

Length (m) per hectare

Once again, $z_i = H_i$ and $f_i(h) = 1$ since $\int_0^{H_i} dh = H_i$. The summand in the estimator, accounting for units, is $[f_i(h)/x_i(h)] = [100/w_i(h)]$. So the estimator for length of logs per hectare is

$$\frac{\widehat{L}}{|\mathcal{A}|} = F_G \sum_i \frac{1}{w_i(h)}$$

Only the ground-width is needed for design-unbiased estimation of log length per hectare.

Surface area (m²) per hectare

Surface area is $s_{a,i} \approx \int_0^{H_i} c_i(h) dh$, where $c_i(h)$ is the relevant portion of the circumference of the log (m). Now the summand in the estimator, accounting for $w_i(h)$ in centimetres, is $[f_i(h)/x_i(h)] = [100 c_i(h)/w_i(h)]$, so the estimator for surface area of logs per hectare is

$$\frac{\widehat{S}}{|\mathcal{A}|} = F_G \sum_i \frac{100 c_i(h)}{w_i(h)}$$

Again, the best method of measuring the ratio $[c_i(h)]/[w_i(h)]$ will depend on the application. But if the log is approximated by a circular cross-section, so that $d_i(h) = w_i(h)$, then $\{[c_i(h)]/[w_i(h)]\} = \{[rw_i(h)/100]/[w_i(h)]\} = \pi/100$, and the estimator becomes

$$\frac{\widehat{S}}{|\mathcal{A}|} = n\pi F_G$$

where n is the number of logs tallied.

Volume (m³) per hectare

The volume, V_i, of a log is, as before, $V_i = \int_0^{H_i} x_{a,i}(h) dh$. Now, $f_i(h) = x_{a,i}(h)$ and the summand in the estimator, $f_i(h)/x_i(h)$, is $100[x_{a,i}(h)/w_i(h)]$. So the estimator for volume of logs per hectare is

$$\frac{\widehat{V}}{|\mathcal{A}|} = F_G \sum_i \frac{100 x_{a,i}(h)}{w_i(h)}$$

Again, the best method of measuring the ratio $x_{a,i}(h)/w_i(h)$ will depend on the application and may depend on the characteristics of the log. If the log is well-approximated by an elliptical cross-section, with one axis oriented vertically and the other horizontally, then

$$100 \frac{x_{a,i}(h)}{w_i(h)} = \frac{100\pi/40\,000\, w(h) d_{v,i}(h)}{w_i(h)}$$
$$= \frac{\pi d_{v,i}(h)}{400}$$

If the log is approximately circular, such that $w_i(h) \approx d_{v,i}(h) \approx d_i(h)$, then

$$\frac{100 x_{a,i}(h)}{w_i(h)} = \frac{\pi w_i(h)}{400}$$

Finally, for highly irregular logs, note that the fraction $x_{a,i}(h)/[w_i(h)/100]$ is just the mean depth of the log (m), measured along $w_i(h)$. As suggested before, for heavily decayed logs, this measurement can be obtained by systematic or random sampling using a sharpened metal rod as a probe.

Log biomass (kg) per hectare

The biomass, b_i, of a log is $b_i = \int_0^{H_i} \rho_i(h) x_{a,i}(h) dh$, where $\rho_i(h)$ is the mean dry density (kg/m³) of an infinitesimal cookie perpendicular to the log axis at h. The fraction $[f_i(h)/x_i(h)] = \{[\rho_i(h) x_{a,i}(h)]/[w_i(h)/100]\}$ can still be well-approximated by measurements on a finitely thin sample cookie cut at h, though now not only the density, but also the dimensions of the cookie are needed. The value of $w_i(h)$ can be measured in the field or on the cookie in the lab if the orientation of the cookie is marked in the field. The cross-section $x_{a,i}(h)$ can be measured accurately using a planimeter or by scanning. As before, if such a cookie is too large for transportation or drying, estimation of the dry mass to volume ratio can be easily turned into a straightforward subsampling problem. The same is true of $x_{a,i}(h)$, for which design-unbiased procedures and accompanying estimators are presented by Gregoire and Valentine (1995).

Practical issues

The estimators above are for single sample points. In practice, of course, one would ordinarily distribute a number of sample points in the tract $\mathcal{A}$. If m sample points are distributed uniformly in $\mathcal{A}$ by simple random or systematic sampling, then the sample mean of the m estimates is a design-unbiased estimate of the population mean. If the points are distributed by simple random sampling then the sample variance, s^2, is a design-unbiased estimate of the variance of the possible sample points and the squared standard error s^2/m is a design-unbiased estimate of the variance of the sample mean. These familiar results can be derived in straightforward fashion from either finite-population Horvitz–Thompson theory (Horvitz and Thompson 1952) or from an infinite-population Monte Carlo approach (Eriksson 1995; Valentine et al. 2001; Williams 2001).

In the field, it is common to encounter logs that are sloped (either because they are elevated above the terrain or because the terrain itself is sloping). Forked and crooked logs also occur. We emphasize that the principles and estimators developed above remain valid in all these situations, but one must be mindful of the geometry of measurement in PDS. In PDS, the log axis is always in the horizontal plane and connects the projections of the two ends of the log onto that plane by a straight line. For forked logs, one of the forks (typically the one with apex farthest from the basal

end of the log) establishes one end. H_i is always measured along this axis and will not equal the physical length of the log when the log is sloping or crooked. All other measurements (such as h, $x_{a,i}(h)$, $c_i(h)$, and $w_i(h)$) are taken perpendicular to this log axis. For sloping logs, this means these measurements remain in a vertical plane and will not be perpendicular to the "physical" axis of the log. For example, cookies to be cut for biomass estimation must be cut in a strictly vertical direction and may slice "on the bias" across the pith of the log. For forked logs, measurements are accumulated among branches in the vertical plane defined by the line of sight perpendicular to the log. For example, $w_i(h)$ might have to be measured as the sum of the ground-width of several forks or branches. Although the estimators presented here are more general, the principles outlined in Williams et al. (2005b) still apply.

Some variables (especially $x_{a,i}(h)$ and $\rho_i(h)$) may be inherently difficult to measure accurately in the field. The impact of this inaccuracy depends on whether these variables appear in the numerator or the denominator of the estimator. If a variable appears only in the numerator and it can be approximated using a design-unbiased sampling method (as is true for $x_{a,i}(h)$ using cookies; Gregoire and Valentine 1995), there will be an increase in variance but no bias. Where such a variable is in the denominator and a design-unbiased estimator is not available or a geometric approximation (e.g., the assumption of an elliptical cross-section) is used, some bias will be present. It may still be possible to correct this bias using an appropriate second-order sampling approach, but bias correction is often counter-productive, owing to the resulting increase in variance (Efron and Tibshirani 1993). We must emphasize that this is a limitation of all known sampling methods for CWM, including fixed-area plots and line intersect sampling. What PDS does avoid with certainty is any bias associated with the assumption of particular taper rates (of diameter, area, volume, or density) along the stem (e.g., Bebber and Thomas 2003).

Finally, we must point out that while the estimators given here are design-unbiased, that does not necessarily mean they will have small variance. For example, the sampling variance for logs per hectare, when using PDS with probability proportional to volume, will almost certainly be very large. (The presence not only of H_i, which may vary considerably from log to log, but also of the random variable $x_{a,i}(h)$, in the denominator of the estimator means these estimates will almost certainly have high variability.) The situation is analogous to that of horizontal point sampling for standing trees; the variance of HPS for basal area is nearly optimal, but the variance for trees per hectare can be atrocious. The solution suggested by Williams et al. (2005a), using PDS to estimate "difficult" variables such as volume or surface area and using a simple count of logs on a small, fixed-area plot to estimate number of logs per hectare, may remain advantageous even when PDS-based design-unbiased estimators for logs per hectare are available. What PDS offers is the ability to sample with probability proportional (or very nearly proportional) to the variable that is of most interest to a particular investigation; that advantage need not be sacrificed simply because more than one variable is of interest.

Conclusions

Previous development of PDS had only allowed for estimating one variable at a time. By focusing on the deep connection between PDS and importance sampling, a general design-unbiased estimator can be developed. The resulting protocols allow simultaneous estimation of multiple CWD variables. In addition to the count of tallied logs, these protocols require only fairly simple and familiar measurements (such as the length of tallied logs or one or more diameter measurements taken at the "perpendicular point" on each tallied log). These results should facilitate the extension of PDS from relatively simple inventories in which only a single variable (such as volume per hectare) is important to more general inventory contexts in which many variables may be of interest.

Acknowledgements

References

Bebber, D.P., and Thomas, S.C. 2003. Prism sweeps for coarse woody debris. Can. J. For. Res. **33**: 1737–1743. doi:10.1139/x03-097.

Brown, J.K. 1974. Handbook for inventorying downed woody material. USDA For. Servi. Intermountain Forest and Range Experiment Station, Gen. Tech. Rep. INT-16.

de Vries, P.G. 1986. Sampling theory for forest inventory. Springer-Verlag, Berlin.

Efron, B., and Tibshirani, R.J. 1993. An introduction to the bootstrap. Chapman and Hall, New York.

Eriksson, M. 1995. Design based approaches to horizontal point sampling. For. Sci. **41**: 890–907.

Gregoire, T.G., and Valentine, H.T. 1995. A sampling strategy to estimate the area and perimeter of irregularly shaped planar regions. For. Sci. **41**: 470–476.

Gregoire, T.G., Valentine, H.T., and Furnival, G.M. 1987. Sampling methods for estimating stem volume and volume increment. For. Ecol. Manage. **21**: 311–324. doi:10.1016/0378-1127(87)90051-X.

Grosenbaugh, L.R. 1958. Point-sampling and line-sampling: probability theory, geometric implications, synthesis. USDA For. Serv. Southern For. Exp. Stn. Occas. Pap. 60.

Hagan, J.M., and Grove, S.L. 1999. Coarse woody debris. J. For. **97**: 6–11.

Harmon, M.E., Franklin, J.F., Swanson, F.J., Sollins, P., Gregory, S.V., Lattin, J.D., Anderson, N.H., Cline, S.P., Aumen, N.G., Sedell, J.R., Lienkamper, G.W., Cromack, K., Jr., and Cummings, K.W. 1986. Ecology of coarse woody debris in temperate ecosystems. Adv. Ecol. Res. **15**: 133–302.

Horvitz, D.G., and Thompson, D.J. 1952. A generalization of sampling without replacement from a finite universe. J. Am. Stat. Assoc. **47**: 663–685. doi:10.2307/2280784.

Ståhl, G., Ringvall, A., and Fridman, J. 2001. Assessment of coarse woody debris — a methodological overview. Ecol. Bull. **49**: 57–70.

Valentine, H.T., Gove, J.H., and Gregoire, T.G. 2001. Monte Carlo approaches to sampling forested tracts with lines or points. Can. J. For. Res. **31**: 1410–1424. doi:10.1139/cjfr-31-8-1410.

von Neumann, J. 1951. Various techniques used in connection with random digits: Monte Carlo methods. Natl. Bur. of Standards Appl. Math. Ser. **12**: 36–38.

Williams, M.S. 2001. New approach to areal sampling in ecological surveys. For. Ecol. Manage. **154**: 11–22. doi:10.1016/S0378-1127(00)00601-0.

Williams, M.S., and Gove, J.H. 2003. Perpendicular distance sampling: an alternative method for sampling downed coarse woody debris. Can. J. For. Res. **33**: 1564–1579. doi:10.1139/x03-056.

Williams, M.S., Ducey, M.J., and Gove, J.H. 2005a. Assessing surface area of coarse woody debris with line intersect and perpendicular distance sampling. Can. J. For. Res. **35**: 949–960. doi:10.1139/x04-203.

Williams, M.S., Gove, J.H., Valentine, H.T., and Ducey, M.J. 2005b. Additional results for perpendicular distance sampling. Can. J. For. Res. **35**: 961–966. doi:10.1139/x05-023.

Appendix A

The design-unbiasedness of the general estimator (eq. 3) can be shown quite simply from a Monte Carlo integration perspective (Valentine et al. 2001; Williams 2001). In this approach, we consider that the attribute of interest for the ith log, z_i, is spread out over the two-dimensional inclusion zone, a_i, of the log. Any attribute that can be expressed as a Lebesgue integral can be z_i

$$z_i = \int_{H_i} f_i(h) \mathrm{d}h$$

where H_i is the axis of the ith log and $f_i(h)\mathrm{d}h$ is the amount of the attribute located in an infinitesimally thin "cookie" of width $\mathrm{d}h$ centered at point h on that axis. H_i, $f_i(h)$, z_i, and a_i are considered fixed for all logs in the population, which is located within a defined tract, $\mathcal{A}$, that has area $|\mathcal{A}|$ (ha).

Denote the sliver of a_i, having vanishingly small width ΔH and crossing perpendicularly at h on H_i, as $a_i(h)$. The length of this sliver perpendicular to the axis is $2kx_i(h)$, as described by the protocol for perpendicular distance sampling (PDS). We spread the attribute z_i within a_i, such that the attribute density within $a_i(h)$ is

$$\rho_i(h) = \lim_{\Delta H \to 0} \frac{f_i(h)\Delta H}{2kx_i(h)\Delta H}$$
$$= \frac{f_i(h)}{2kx_i(h)}$$

at any and all points in $a_i(h)$. For all points outside a_i, $\rho_i(h)$ is defined as zero. Note that the attribute density $\rho_i(h)$ is not to be confused with physical density.

Now, suppose we select a sample point $s \in \mathcal{A}$, with probability density $p(s)$. Denote $h_i(s)$ as the value of h for the ith log when the sample point is s. (When no perpendicular point exists, the definition of $h_i(s)$ is irrelevant.) Then, an unbiased estimator of $Z = \sum_{i=1}^{N} z_i$ is

$$\widehat{Z} = \sum_{i:s\in a_i[h_i(s)]} \frac{\rho_i[h_i(s)]}{p(s)}$$

If s is selected uniformly at random within $\mathcal{A}$ then $p(s) = 1/|\mathcal{A}|$ and

$$\widehat{Z} = \frac{|\mathcal{A}|}{2k} \sum_{i:s\in a_i[h_i(s)]} \frac{f_i[h_i(s)]}{x_i[h_i(s)]}$$

Note that in the above expression, $|\mathcal{A}|$ and $|a_i|$ are assumed to be expressed in the same units. If that were not the case (e.g., $|\mathcal{A}|$ is in hectares and $|a_i|$ is in square metres), an appropriate unit conversion would be needed. The corresponding estimator of Z per unit area is

$$\frac{\widehat{Z}}{|\mathcal{A}|} = \frac{1}{2k} \sum_{i:s\in a_i[h_i(s)]} \frac{f_i[h_i(s)]}{x_i[h_i(s)]}$$

In the special case in which $z_i = \int_{H_i} x_i(h)\mathrm{d}h$, for example when z_i is volume and $x_i(h)$ is cross-sectional area,

$$\frac{\widehat{Z}}{|\mathcal{A}|} = \frac{1}{2k} \sum_{i:s\in a_i[h_i(s)]} \frac{x_i[h_i(s)]}{x_i[h_i(s)]}$$
$$= \frac{n(s)}{2k}$$

where $n(s)$ is the number of logs tallied from point s. This is the estimator developed by Williams and Gove (2003).

The proof of design-unbiasedness is straightforward. By our Monte Carlo design

$$\rho_i(s) = \begin{cases} \rho_i[h_i(s)] & s \in a_i; \\ 0 & s \notin a_i \end{cases}$$

Hence,

$$z_i = \int_{s\in a_i} \rho_i(s)\mathrm{d}s$$
$$= \int_{s\in \mathcal{A}} \rho_i(s)\mathrm{d}s$$

Therefore, taking $p(s) = 1/|\mathcal{A}|$ and assuming no slopover

$$E\left[\frac{\widehat{Z}}{|\mathcal{A}|}\right] = \int_{s\in\mathcal{A}} p(s)\sum_{i=1}^{N}\rho_i(s)\mathrm{d}s$$
$$= \frac{1}{|\mathcal{A}|}\int_{s\in\mathcal{A}}\sum_{i=1}^{N}\rho_i(s)\mathrm{d}s$$
$$= \frac{1}{|\mathcal{A}|}\sum_{i=1}^{N}z_i$$
$$= \frac{Z}{|\mathcal{A}|}$$

References

Valentine, H.T., Gove, J.H., and Gregoire, T.G. 2001. Monte Carlo approaches to sampling forested tracts with lines or points. Can. J. For. Res. **31**: 1410–1424. doi:10.1139/cjfr-31-8-1410.

Williams, M.S. 2001. New approach to areal sampling in ecological surveys. For. Ecol. Manage. **154**: 11–22. doi:10.1016/S0378-1127(00)00601-0.

Williams, M.S., and Gove, J.H. 2003. Perpendicular distance sampling: An alternative method for sampling downed coarse woody debris. Can. J. For. Res. **33**: 1564–1579. doi:10.1139/x03-056.